

CALIFORNIA UNIV LOS ANGELES DEPT OF PSYCHOLOGY
PUPIL DILATION AS AN INDEX OF WORKLOAD.(U)
FEB 80 J BEATTY

N00014-76-C-0616

TR-20

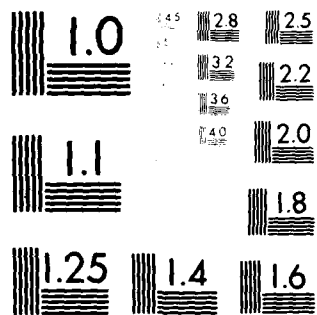
NL

1961
 1962

END

GATE

FILMED
9-80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL

12
P.S.

TECHNICAL REPORT #20

PUPIL DILATION AS AN INDEX OF WORKLOAD

JACKSON BEATTY
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF CALIFORNIA AT LOS ANGELES

1 FEBRUARY 1980

DTIC
ELECTE
AUG 1 8 1980
S D C

PREPARED FOR OFFICE OF NAVAL RESEARCH
PHYSIOLOGY PROGRAM, ENVIRONMENTAL PHYSIOLOGY
CONTRACT N00014-76-C-0816

REPRODUCTION IN WHOLE OR IN PART IS PERMITTED FOR ANY
PURPOSE OF THE UNITED STATES GOVERNMENT

This document has been approved
for public release and sale; its
distribution is unlimited.

80 3 15 130

AD A088022

DDC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 20	2. GOVT ACCESSION NO. AD-A088 022	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PUPIL DILATION AS AN INDEX OF WORKLOAD.		5. TYPE OF REPORT & PERIOD COVERED Interim Technical Report
7. AUTHOR(s) 10 Jackson Beatty		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of California Los Angeles, California 90024		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0616
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 441) Department of the Navy Arlington, Virginia 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 201 - 207
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) UTR-1R		12. REPORT DATE 1 February 1980
		13. NUMBER OF PAGES 16
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Presented at the 1978 meeting of the Air Line Pilots Association.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) pupillometrics cognitive processing mental workload		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The results of a recent series of experiments on task-evoked pupillary responses are reviewed. These data form the basis for a physiological theory of processing resources and mental workload that has practical implications for pilot workload measurement.		

Pupil Dilation as an Index of Workload

**Dr. Jackson Beatty
University of California**

I would like to describe some work that has been going on in our laboratory and in some other laboratories for a number of years. The work in our own laboratory has been sponsored by the Advanced Research Projects Agency of the Department of Defense and by the Office of Naval Research.

This work has addressed the problem of measuring the central processing component of workload in the terms introduced this morning by Dr. Wickens. We are very comfortable with his distinction between the cognitive and other types of capacity. We have sought to measure cognitive workload physiologically in a variety of tasks. Our approach differs from that described by other speakers in two respects. One is that we are concerned with the direct response of the nervous system to increases or decreases in processing load rather than responses measured to probes that are extraneous to the ongoing task. The second difference is that we are concerned with cortical-subcortical relationships. What we are really interested in is not so much the activity of the cortex, which you have heard a great deal about today, but rather the activity of the brain stem, which supports cortical information processing.

The brain stem lies beneath the cerebral cortex. It is essentially the upper projection of the spinal cord that becomes increasingly more complex and differentiated at its higher levels.

Now during human information processing, the cortex is put under processing load by cognitive tasks. The cortex then sends requests down to the brain stem seeking additional activation to support the increased processing load. Running through the center of the brain stem is a massive structure called the reticular formation. The response of the reticular formation to demands for increasing load is a forward surge of -- well, let us just call it activation, and pass by its intricacies for the moment.

Now, if we think that is what goes on during cognitive processing, and if we think that the activity of that brain stem activating system is intimately related to processing load, we are faced with the problem of how we are going to measure activity in such an inaccessible region of the brain. The answer to that question is to

exploit anatomical coincidence. The brain stem also contains a set of intricate pathways that control some of the autonomic reflexes, particularly the movements of the pupil of the eye.

Because this brain stem region is relatively undifferentiated, it appears to be possible to measure changes in brain stem reticular formation activity in response to central or cortical demands by looking at very small changes in the dynamics of pupillary movement, which are regulated at the same level of the brain. It is a convoluted way of approaching this problem, but one that we think has some interesting features.

The pupil of the eye is the hole in the center surrounded by the muscular iris. There are two sorts of fibers in the iris. One group is dilator fibers, which are sympathetically innervated. When they are activated, they act to open up the pupil. The other set is sphincter muscles. They are parasympathetically innervated, and when they are activated, they act to close the pupil down.

Now brain stem activation has been shown to affect pupillary movements. In experiments with animals, artificial electrical stimulation of the reticular formation produces cortical signs of activation.

The pupillary effects of reticular formation are far easier to observe. In the parasympathetic pathway to the sphincter, reticular stimulation produced inhibition. In the sympathetic pathway to the dilator muscles, reticular stimulation greatly increases activation. All this anatomical detail is intended to do is to make plausible the long-known fact that stimulation of the brain stem reticular activating system results in pupillary dilation. This is how it does it. It is not just happenstance.

With that as a background, then, we are going to attempt to exploit anatomical fact to look at the brain stem's responses to changes in processing load. To do this in our laboratory we use a Whitaker Pupillometer. The Whitaker Pupillometer is essentially a signal processing unit that looks at the image of the eye obtained by a high-resolution infrared video camera and extracts a measure of pupillary diameter. I am not going to go into how all that works, except to say that it is a remarkably reliable and sensitive instrument, and it exists in a couple of other forms besides the laboratory version. For example, there is a more complex model that allows accurate pupillary measurements to be made over a range of some degree of head movement, by using another camera to track the head.

What is the evidence that these sorts of pupillary movements might tell us anything at all about the physiological basis of workload assessment? The answer comes from a long series of behavioral studies. This first study that I am going to show you now was published 10 years ago by Danny Kahneman and myself. What we looked at was a classical task that imposes cognitive workload. It is a short-term memory task, in which people are asked to remember strings of numbers read to them at the rate of one per second. Strings ranged

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DDC TAB	<input type="checkbox"/>
unannounced	
notification	
Distribution/	
Availability Codes	
st	Avail and/or special
A	

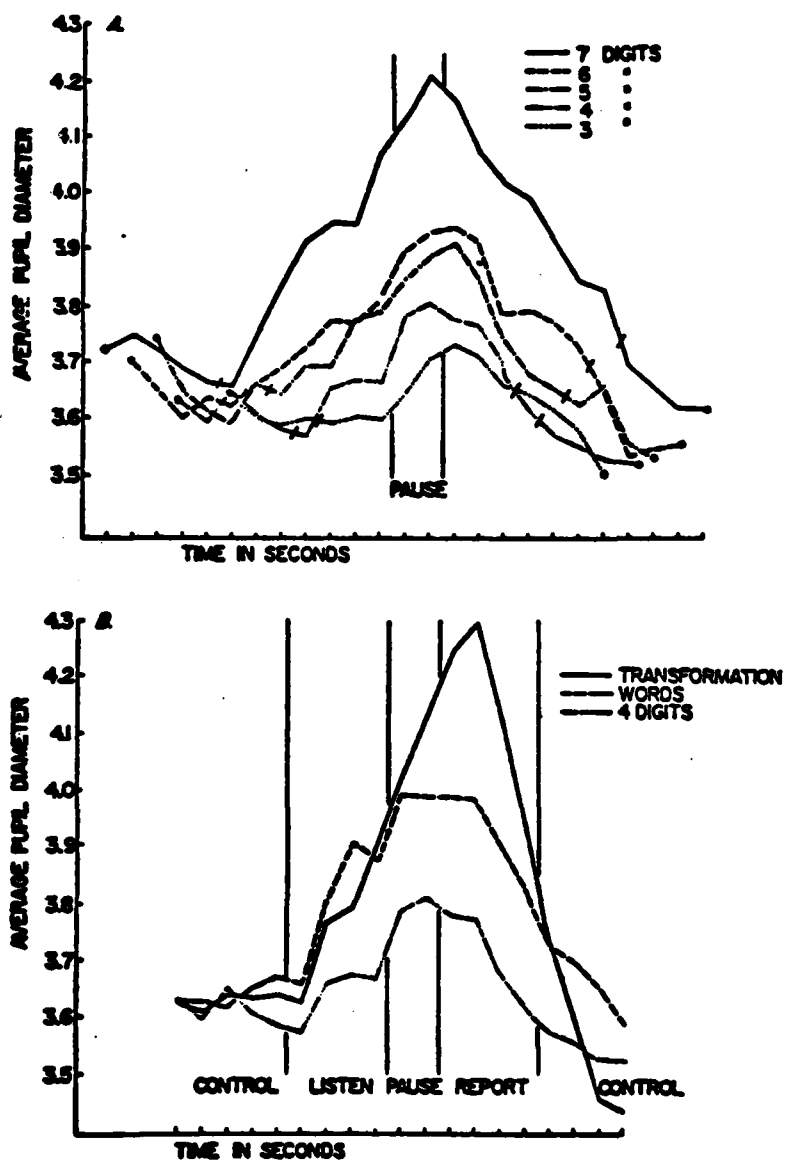


Figure 1. Upper graph: Average pupillary diameter during presentation and recall of strings of 3 to 7 digits, superimposed about the two second pause between presentation and recall. Slashes indicate the beginning and the end of the memory task. Lower graph: Pupillary diameter during presentation and recall of four digits, words and a digit transformation task.

in length from three to seven digits. The task was to repeat back the number accurately after a two-second pause.

The average evoked pupillary responses during the performance of this digit span task for strings of three, four, five, six and seven digits are shown in the upper portion of figure 1. These are absolute pupillary diameters. Because the tasks are of different length, they are all superimposed above the two-second pause in the middle. The thing that should be immediately apparent from figure 1 is that during the loading of short-term memory there is an increase in pupillary diameter, as each digit is heard and encoded, that reached a maximum in the pause between encoding and report. As the items are reported back from memory, the pupil returns to its baseline level.

The second thing you should note is that the increase in pupillary diameter at the pause is a function of the number of digits stored, and it increases systematically to seven digits. On the face of it, then, this appears to index some aspects of cognitive processing that should be related to load.

The lower part of figure 1 shows that it is not just the number of items but also the sort of items that makes a difference. All three curves are for four items. In the lower curve the subject is listening to four digits and saying them back. The middle curve is the same task but for four unrelated words. Memory span is shorter for four unrelated words, therefore one might assume that the load imposed by each word is greater. And, in fact the slope of the pupillary function can then be seen to be greater for the more difficult items.

The third task is a mental transformation task in which the person hears four digits, adds one to each one of them and reports back a transformed string. You can see that the transformation task results in the largest pupillary dilation.

Now, how do we know that has anything to do with workload, outside of the fact that it seems somewhat plausible? In a second experiment, Kahneman, Pollack and I essentially verified the workload changes using the standard two-task measurement technique, in which the second was a visual scanning task. You do not need to worry too much about the details of this experiment. Figure 2 shows performance on the secondary task, which is scanning a visual display for a letter K that may occur at one of five positions, during the simultaneous performance of the memory task and in its absence. You can see that secondary task performance deteriorates as processing load increases, a standard sort of behavioral interference result.

Figure 3 shows some data from Scott Peavler's laboratory at Bell Telephone Laboratories that have some very interesting properties. What he was looking at is pupillary response for strings of items that are both under and over behavioral capacity. You can see that the first digit is presented at the third second, and you see a pupillary dilation that goes up until about the seventh item is presented. From there on out it levels off. As is well known in human performance psychology, the digit span is in fact about seven items. What this

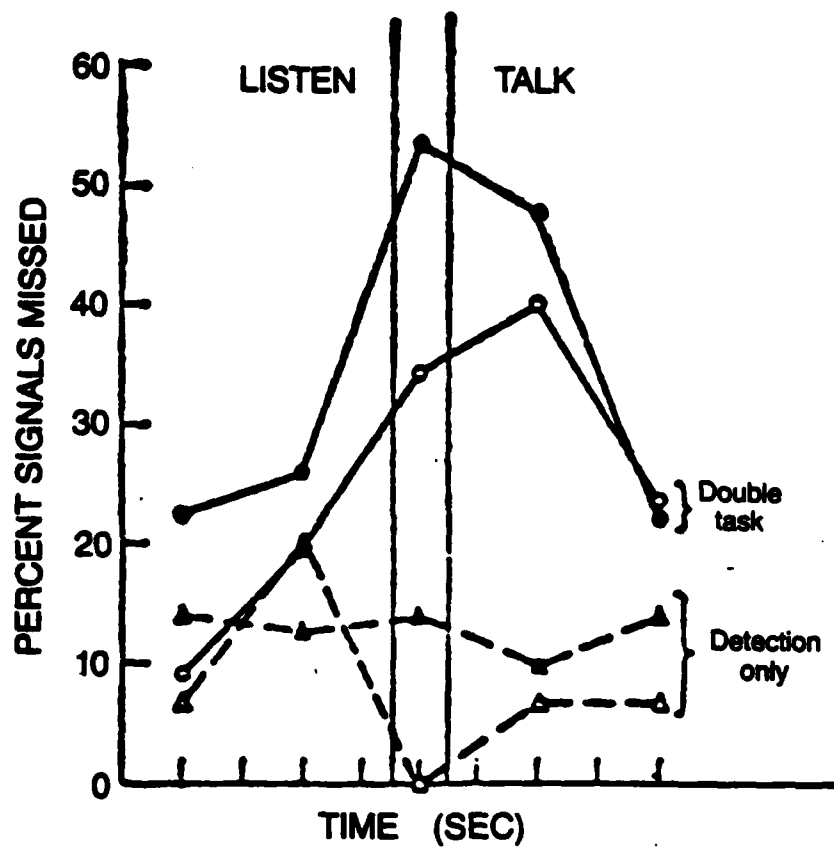


Figure 2. Behavioral evidence confirms the use of pupil as a measurement of workload. In the same short-term memory task, when performed with a simultaneous detection task, the pupillary data predict perfectly the pattern of behavioral interference.

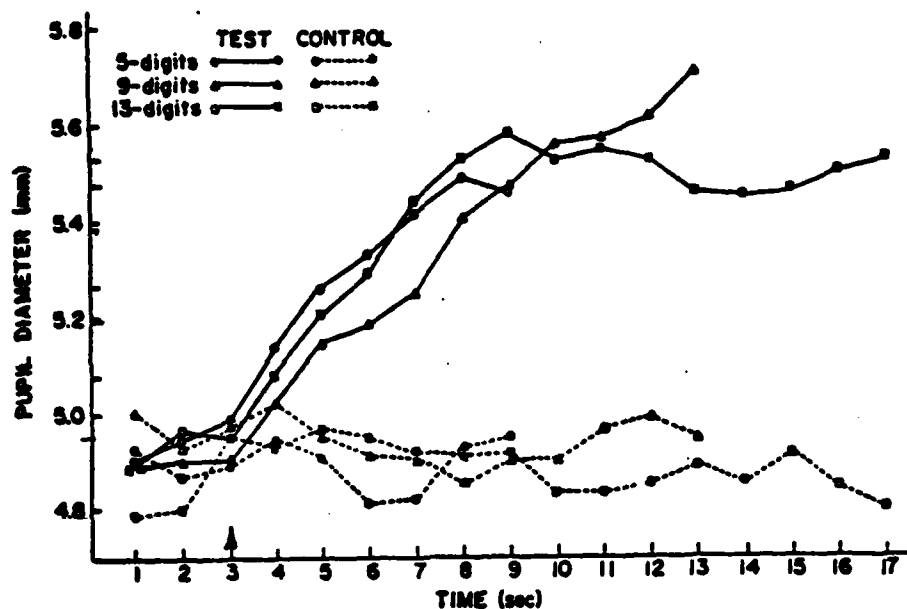


Figure 3. Averaged pupillary diameter for 14 subjects during presentation of strings of 5, 9 and 13 digits for immediate recall (solid lines). The broken lines are taken from non-processing control trials. Filled symbols represent pupil size during digit presentation and unfilled symbols represent pupil size for 2 seconds preceding and following presentation of the digit strings. Notice that the pupillary functions asymptote at approximately 7 items (about second 10 on the abscissa).

says is that at the peak of capacity, putting additional information into the person for processing no longer results in increased physiological response. So that the physiological response to increased processing load saturates at the behavioral limit of capacity.

What I am going to do now is to show you a few other sorts of cognitive tasks. I hope to convince you that this kind of autonomic response to processing load is not somehow uniquely associated with the digit memory task. Figure 4 presents results from a pitch discrimination task, in which a standard tone of 850 hertz is first presented. Four seconds later a comparison tone is presented, and the person's task is to say whether the comparison tone is higher or lower in pitch.

In the case of an easy discrimination -- here it is 30 hertz difference, that is the 880 curve -- you see that the dilation to the comparison tone is small compared with the dilation for an impossible discrimination, in which the comparison tone and the standard are the same pitch. A difficult discrimination yields a large response.

Figure 5 shows us the magnitude of the dilation to the comparison tone as a function of the difficulty of the discrimination. The dotted curve gives the amplitude of the pupil response to the comparison stimulus. The solid curve gives some behavioral data, which is the percentage of errors that occur for comparison stimuli for those particular frequencies. You can see that there is a reasonable correspondence between those curves.

Figure 6 presents some recent data obtained in our laboratory by Sylvia Ahern as part of her doctoral dissertation. This is a mental multiplication task, a more cognitive sort of task than the ones that we have described. People are presented with problems of either multiplying one digit by one digit, one digit by two digits or two pairs of two-digit numbers together. Pupillary dilations are present in the average evoked pupillary responses, first as the multiplicand is presented and second as the multiplier is presented and the problem solved. You can see that both the duration and the peak magnitude of the response vary as a function of presumed cognitive load of the mental arithmetic task.

Figure 7 presents the results of another experiment from Ahern's dissertation, one that involves the comprehension of linguistic information in a test devised by Alan Baddeley. Here, people were presented with a sentence of the form "A precedes B," "A is followed by B," "A is not followed by B," "A is not preceded by B," et cetera. After the presentation of each sentence, which was done auditorily, the letter pair AB or BA was presented. The person's task was to judge whether the sentence correctly described the ordering of letters in the letter pair.

Two things are apparent in these data. First is the general form of the pupillary response that is characteristic of this sort of task, a long, slow dilation. Secondly, there are small differences between

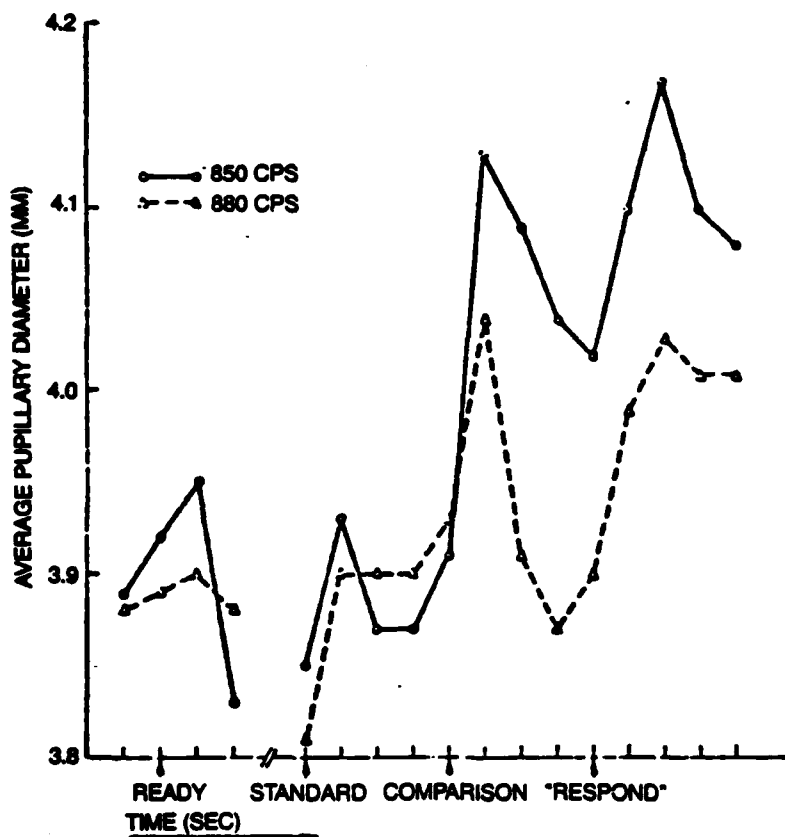


Figure 4. Average pupillary diameter for 10 subjects (5 trials per subject) for two discrimination tasks. The frequency of the comparison tone is the parameter. Standard tone frequency is always 850 cps.

the amplitudes of these responses that depend upon sentence type, with the more complex, transformationally more difficult sentences eliciting greater activation.

This is a sentence comprehension type of task that is quite different from the mental multiplication problem, the sensory discrimination test or the short-term memory task.

Now, one thing that you might have noticed, if you were a very careful watcher of these illustrations is that they were arranged so that the curve would occupy a reasonable proportion of the chart. For that reason, the magnitude or the size of the scaling differed between experiments. It turns out that there are some very interesting properties to this pupillary response. One is that the magnitude of the task-evoked pupillary response is independent of the baseline pupil size. This was first shown by Bradshaw in 1969, and it holds over a wide but not extreme range of basal pupil sizes.

What that means is that one can go through the experimental literature and directly compare the results obtained for different types of cognitive tasks by comparing the magnitude of the average evoked pupillary response, and come up with a scale that gives some indication of the magnitude of the brain stem response to qualitatively different tasks.

Figure 8 shows the result of such a compilation. Memory for immediate digits, the digit span, I have taken as sort of a standard. These data come from the work of Scott Peavler. It could just as well be mine or the earlier work with Kahneman. All of those values line up quite closely. It is across approximately ten years of work, that this sizing scale holds.

The right-hand column of figure 8 shows the peak pupillary dilation to other cognitive tasks: two by two digit multiplication, two by one digit multiplication, one by one digit multiplication, the Baddeley tasks, difficult sensory discrimination and easy sensory discrimination. The smallest dilation was observed in a study we published recently in Science. It involved just looking at a pair of letters and saying whether they were the same or different, a very, very easy sort of task with a very small pupillary response.

I would like to point out two things. One is that the average-evoked pupillary response seems to constitute a reasonable physiological measure of load. The other is that in reviewing the literature I came across an interesting fact in some of Alan Baddeley's work. He used his sentence comprehension task -- this is "A precedes B, AB" -- and examined the effects of concurrently doing a short-term memory task. He would give people some numbers to remember, then give them the sentence to judge whether it is correct or not, ask for the judgment on the sentence, and then ask for the numbers back.

What Baddeley reported was that if a person had to remember one or two digits, there was no interference with performance on the sentence recognition task. If he has to remember more than that, there is

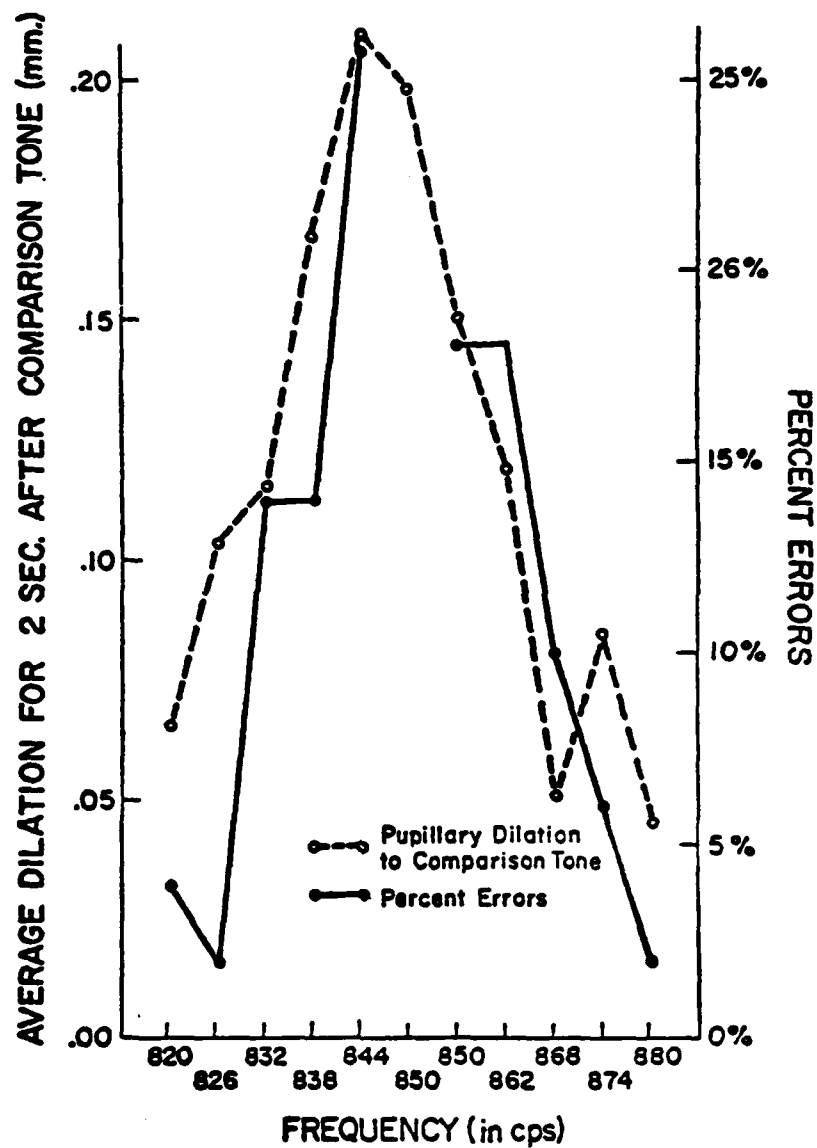


Figure 5. Average pupillary dilation during the decision period and the percent errors in a pitch discrimination task as a function of the frequency of the comparison tone. The frequency of the standard was 850 Hz.

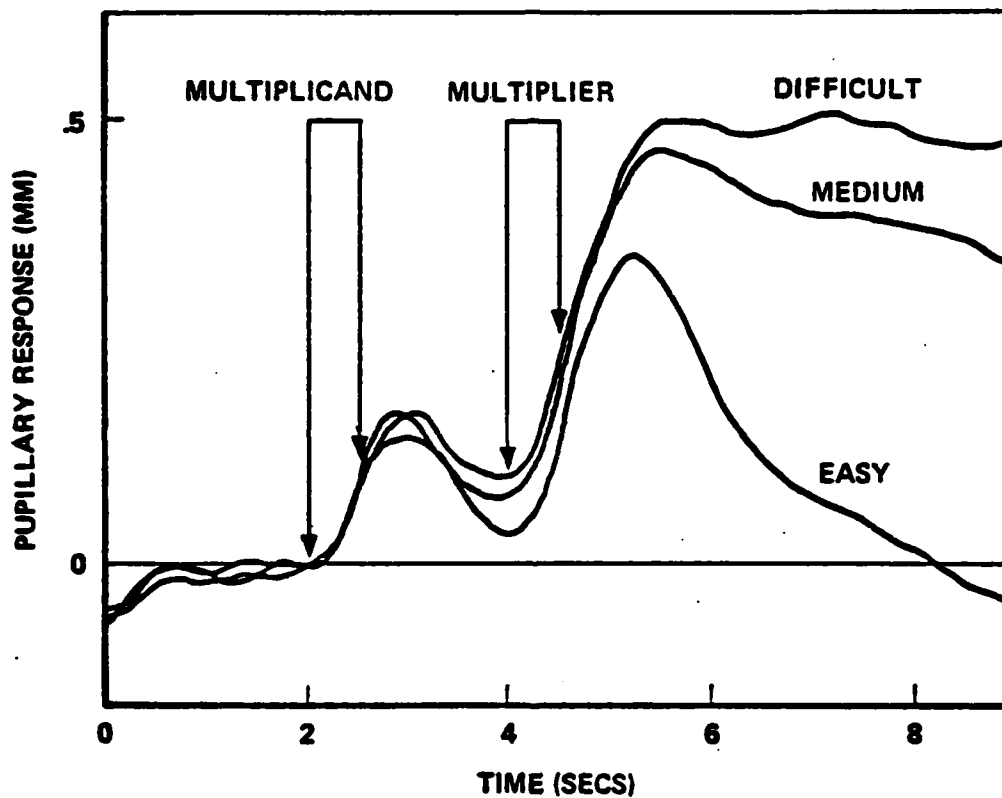


Figure 6. Averaged evoked pupillary responses in a mental multiplication task. Two dilations are evident: one following presentation of the multiplicand and a second following the multiplier. The amplitude and duration of the pupillary response increases with problem difficulty.

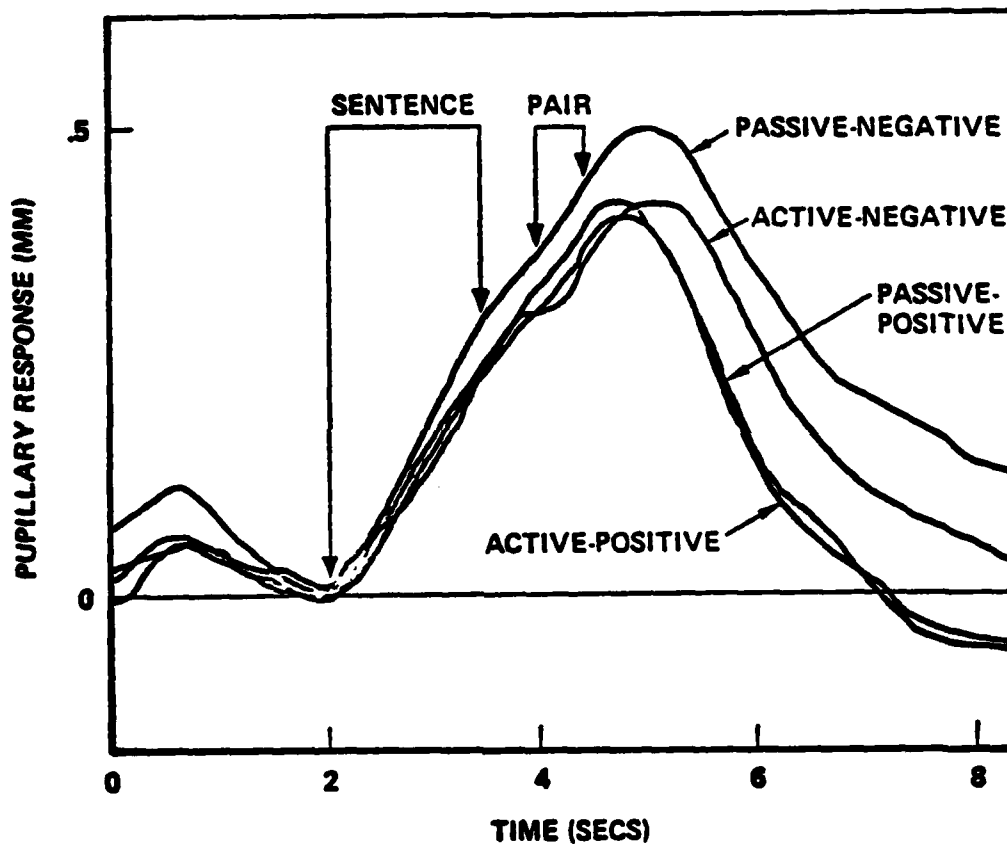


Figure 7. Averaged evoked pupillary responses in a grammatical reasoning task. The smallest dilations are observed in the processing of simple active sentences and the largest with passive negative sentences. These data suggest that sentences of varying transformational complexity differ also in processing load.

interference. In other words, the short-term memory task for more than two items interfered with the performance of the grammatical reasoning task. What is interesting is that if you look at the average level of the pupillary response for the sentence comprehension task and if you were to accept that it is a scale of physiological load, then you would expect approximately two or three more items would be needed before interference should occur.

Where does that leave us? All of this is discussing cognitive tasks, albeit perhaps a wide range of cognitive tasks, that can be performed in the laboratory. What bearing does this have on the question of measuring pilot workload? I would suggest that there are at least three things that one might be interested in.

One might be interested in a discrete task analysis, in which various components of the cockpit behavior could be pupillometrically measured in the laboratory situation and a scale of physiological load for these components could be constructed. From that ordering of various component processes, one might be able to make inferences as to what things can be combined safely and what things may not.

Second, this sort of procedure seems to offer the possibility of a quantitative measure of the physiological load imposed by an information-processing task that bears at least some surface resemblance to the classical question of workload. Given that assumption, then one could look at the questions of fatigue, for example, by looking at the load imposed upon the nervous system under rested and under fatigued conditions.

Third, one could approach the problem of jet lag or time-zone crossing in a similar manner and ask whether the physiological load imposed by any sort of an information-processing task is greater as a function of lag.

These would be the directions in which I believe that the physiological analysis of cognitive workload or processing load might have some bearing on the question of pilot workload and pilot performance.

MS. HART: Sandra Hart, Ames Research Center. If you are trying to measure the impact that different sorts of displays might have on the pilot's workload, how would you use pupil size to measure this? Because as you added different elements to the displays, the lighting will vary from moment to moment. Could one use this sort of measure under those conditions?

DR. BEATTY: This is something we have thought about at some length, and I guess my answer to that is that you have to be really careful. We have indeed used visual displays in our experiments, but we have had to take special pains to make sure that they are equiluminous displays in time. Also, we have found low-contrast displays to be useful. Now, if you do these and are careful about it, you can get around working with visual displays.

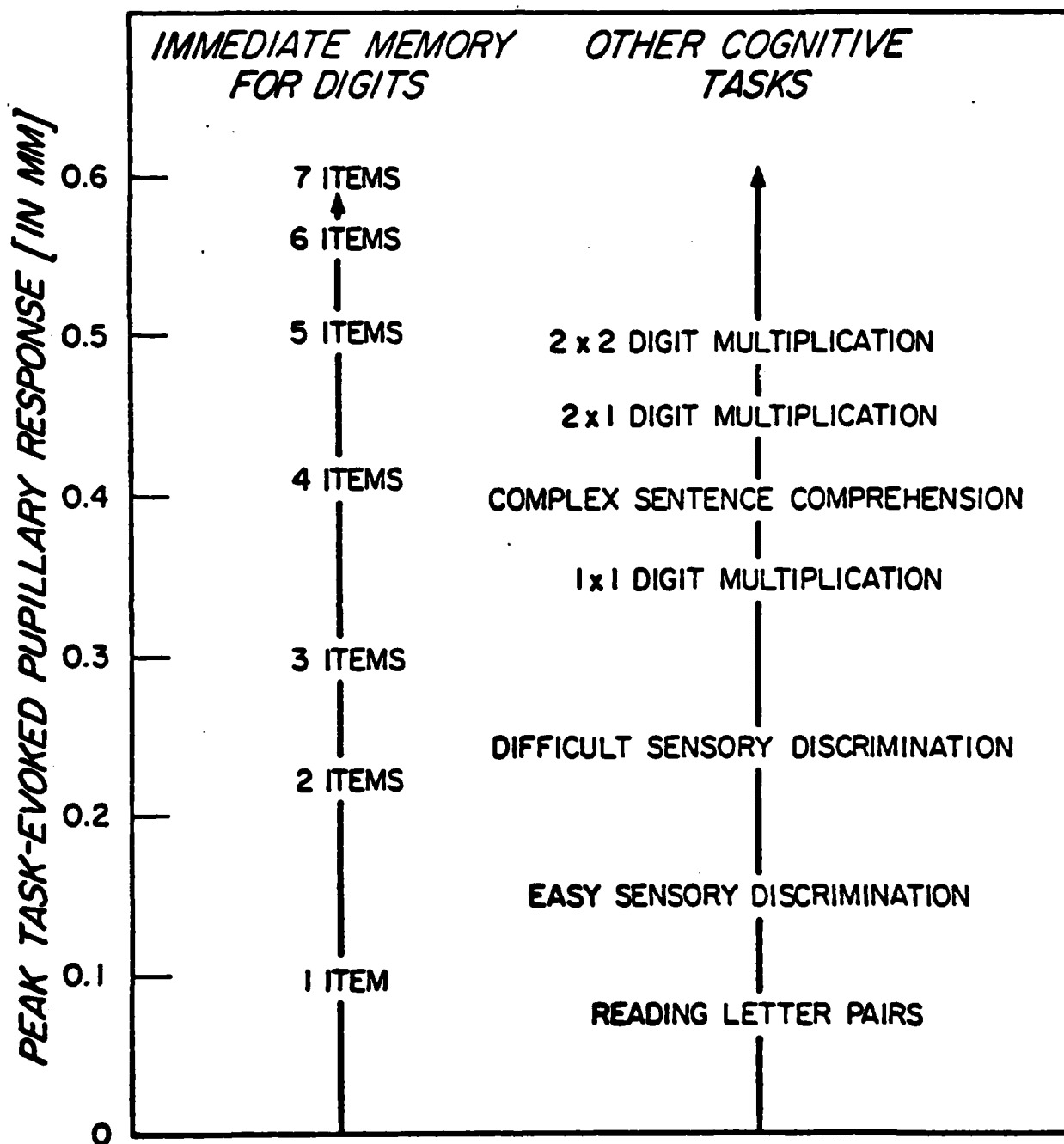


Figure 8. Magnitude of the peak task-evoked pupillary responses for a variety of qualitatively different mental tasks. The ordering suggests the feasibility of physiological measurement of processing load.

But the problem is a severe one, and that is why I would suggest that the first step should be discrete task analysis, even with visual displays, in a laboratory situation rather than going to the free-ranging pupil in a cockpit, with sunlight coming in one window and dark displays down below.

MR. WATT: Bill Watt, Executive Air Fleet. What difference does IQ of the subject make?

DR. BEATTY: Sylvia Ahern's dissertation, the one from which you saw the mental multiplication and grammatical reasoning tests, shows one of the first reliable physiological differences between people of high and normal IQ. Specifically, on the more complex information processing tasks such as mental multiplication, the brighter person can perform a task at a lower level of loading than can his less-gifted counterpart. The pupillary response curves for the brighter people reach lower peaks for solving the identical problem than do the responses from not-so-bright folk.

There is another part of the dissertation that was not adequately tested because of some procedural difficulties. We wanted to also look at the question of whether, when pushed to the limit, bright people have more of a reserve of activation than do not-so-bright people, and we just do not know the answer to that question.

DR. GERATHEWOHL: I am Dr. Gerathewohl, FAA. Your curves were always showing a baseline of about eight seconds. What happens after this, at an extended period of time? Do your curves still stay discriminated, or do they go back to baseline pattern, and if so, at what time?

DR. BEATTY: The reason for the 8-second durations here is because that is the length of the tasks. When the task is over, the pupil returns to its baseline level. However, for longer active tasks, the pupil continues to reflect momentary load. When a series of short-term memory tests are given over a period of an hour, for example, the pupil remains as accurate an indicator of load at the end as it was in the beginning.

CAPT. MOUDEN: Homer Mouden, Flight Safety Foundation. What impact does the stress level have on the pupillary response time, and what is the shortest interval of response time that you have been able to identify?

DR. BEATTY: Real stress levels, I do not know. I can only cite two things. One is the general finding that baseline pupil diameter does not interact with the magnitude of the phasic response, over a wide range of conditions. The second is a study in which amphetamines were given. I think it was a simple reaction-time task. Amphetamines changed baseline pupil size but did not change either reaction time or the magnitude of the task-evoked pupillary response.

In answer to your second question, in an acoustic signal detection task, where weak signals occurred at unspecified times against a background of white noise, we would see a task-evoked pupillary response

to the occurrence of a signal that would have a time from latency to peak of about 600 milliseconds. But the time from latency to first reliable differentiation between that and the case in which there is no response is on the order of 100 to 200 milliseconds. The peak is slow, but the differentiation occurs in less than a quarter of a second.

